JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN EXTENSIVE RESEARCH ON BIG DATA CLOUD DRIVEN COMPUTATIONAL FRAMEWORK FOR INTELLIGENT TRANSPORTATION SYSTEM

D. Venkata Siva Reddy[1], Dr. M. Senthil Kumaran[2]
1. Research Scholar, Department of CSE, SCSVMV University, Enathur, Kanchipuram. Email-id: lionshivareddy@gmail.com
2. Associate Peofessor & Head, Department of CSE, SCSVMV University, Enathur, Kanchipuram.
Email-id:senth.kumaran2000@gmail.com

Abstract

The concept of big Data for intelligent transportation system has been employed for traffic management on dealing with dynamic traffic environments. Big data analytics helps to cope with large amount of storage and computing resources required to use mass traffic data effectively. However these traditional solutions brings us unprecedented opportunities to manage transportation data but it is inefficient for building the next-generation intelligent transportation systems as Traffic data exploring in velocity and volume on various characteristics.

In this work, a new framework named Big Data Cloud Driven Computational Framework for Intelligent Transportation System has been architected on employing a new architectures named as deep intelligent network, Increase weighted continuous feature extraction and reinforcement learning. These architectures has been introduced to handle the magnifying and exploring data with hierarchical and spatiotemporal characteristics along location based service by utilizing the Sensor and GPS data of the vehicle in the real time. The proposed framework employs deep learning architecture to predict potential road clusters for passengers on analysing the various characteristics on the attributes of the dataset. In order to perform efficient big data analysis, Feature extraction model has to be structured properly in order to identify high scaling features. Initial methodology of the work is to construct the Increased weighted Continuous feature extraction Algorithm which includes filter, wrapper and hybrid method on efficient feature extraction for classification of road clusters using support vector machine.

Reinforcement learning model has been modelled as second architecture for predicting the salient feature on the time varying traffic contexts by approximating non linear data in the dataset. It is injected as recommendation system to passenger in terms of mobile application by employing the agent and feedback function through iterating optimum control policy to control the hardware equipment employment on the vehicle incorporating location based services models to seek available parking slots, traffic free roads and shortest path for reach destination and other services in the specified path etc. The underlying the traffic data is classified into clusters with extracting set of features on it.

Finally, deep behavioural network has been architected to processes the traffic data in terms of spatiotemporal characteristics to generate road clusters to produce the traffic forecasting information, vehicle detection, autonomous driving and driving behaviours. In addition, markov model is embedded to discover the hidden features .The experimental results demonstrates that proposed framework achieves better results on the performance measures named as precision, execution time, feasibility and efficiency.

Keywords - Big Data, Intelligent Transportation System, Deep Learning, Prediction, Spatiotemporal analysis, Location based services.

1. Introduction

Big data solutions have been employed in all areas of research perspectives and it is becoming undeniable. In particular, big data techniques have become more popular in health care and transportation system [1]. Especially transportation has much attention due heavy exploration and utilization of vehicles leading to large traffic congestions and other incidents. Due to large exploration of traffic data, it becomes mandatory to model architecture to predict and cluster the traffic data based on the various constraints and dynamic traffic environment.

Normally traffic data is collected using GPS devices, Sensors and probes. The unsupervised learning model has been applied in large extent to classifying and clustering of the traffic data. Feature selection method which is also termed as attribute or variable subset selection method is used in opting subset with appropriate predictors or features. Refined traffic model has helped in extending reinforcement learning with new representative and reward functions on influencing characteristics of the data.

Prediction model have shown high importance in recommendation and forecasting services in an effective manner. Thus underlying traffic information is processed with deep learning models for clustering based on set of extracted features [4]. The deep learning of the traffic data predicts the potential road clusters for passengers, traffic forecasting information, Vehicle detection and helps in autonomous driving behaviours.

Due to the data diversity, there are several hot branch points for each topic such as aggregated and disaggregated model. In intelligent transportation systems, different frameworks for big data analytics on transportation management and operations are introduced by using Hadoop to analyse the sensor traffic information, road cluster and travel patterns on basis of location based service and spatial temporal characteristics from the distributed servers.

Statement of problem

Intelligent Transportation system is an accurate and timely management of the traffic flow information which currently strongly needed for individual travelers, business sectors, and government agencies. It has the potential to help road users make better travel decisions, alleviate traffic congestion, reduce carbon emissions, and improve traffic operation efficiency. Traffic flow prediction heavily depends on historical and real-time traffic data collected from various sensor sources, including inductive loops, radars, cameras, mobile Global Positioning System, crowd sourcing, social media, etc. The extensive routine traffic volumes bring pressures to existing urban Infrastructures resulting in the traffic congestion.

With the widespread traditional traffic sensors and emerging traffic sensor technologies, traffic data are exploding and transportation management and control is now becoming more data driven. Although there have been already many traffic flow prediction systems and models using machine learning paradigm. It functions as through supervised and unsupervised learning which makes decision based on the output labels.

Intelligent transport systems (ITS) has been radically transforming to automatic by the emergence of the big data streams generated by the Internet of Things (IoT), smart sensors, surveillance feeds, social media, as well as growing infrastructure needs. It is timely and pertinent that ITS harness the potential of an artificial intelligence (AI) to develop the big data-driven smart traffic management solutions for effective decision-making. The existing AI techniques that function in isolation exhibit clear limitations in developing a comprehensive platform due to the dynamicity of big data streams, high-frequency unlabeled data generation from the heterogeneous data sources, and volatility of traffic conditions.

Machine learning algorithms lags on many aspects as it use shallow traffic models time saving for drivers, energy saving for environment, and safety for all participants has been analyzed on the following aspects.

- Learning architecture employed for processing the large scale traffic data has failed to incorporate the changes in the data velocity and volume.
- Learning models suffers from high computational complexities and it neglects the utilization of the influencing characteristics of the data
- Leads to curse of dimensionality and sparsity issues due to exploration of dimension of the large scale dataset.
- It is complicate to calibrate and validate the traffic information with heterogeneity in distributed sources.
- Data consistency as data conciliation issues due to infinite state problem.

Objective of the work

In order to derive valuable insights for decision support and automation from huge volumes of heterogeneous sensor data, a few of big data technologies have been focused and invested using deep learning architecture using three categories, including data storage (such as, Hadoop, Data Lakes, NoSQL Databases, etc.), data processing (such as, Spark, Hadoop, data governance, etc.), and data analytics (Spark, cloud computing, edge computing, artificial intelligence, modeling and optimization, etc.). Specifically, the Hadoop ecosystem has been widely recognized due to its reliable, efficient and scalable distributed processing of large dataset.

The major objective of the research is

- An efficient and scalable system for storage and data management has to be envisioned as intelligent prediction model for volume and velocity changes in the transportation data through distributed function to processing the high dimensional and infinite state problem effective application of data reduction techniques and feature extraction techniques.

- Location based service has to be incorporated in the data prediction and classification architecture to suggest the available parking information, traffic density on the specified location along the multisource data in the data driven applications without any prior knowledge. It is eliminates the computational complexities using hyper parameter tuning on the feature selection

- Reinforcement learning has to be employed to determine the idea behaviour within specific content of the location data processes the traffic data in terms of spatiotemporal characteristics to generate the traffic forecasting information, vehicle detection, autonomous driving and driving behaviours on incorporating hidden features using hidden markov model.

2. RESEARCH METHOLOGY

The methodology of the work towards establishing an autonomous intelligent transportation system is structured as follows

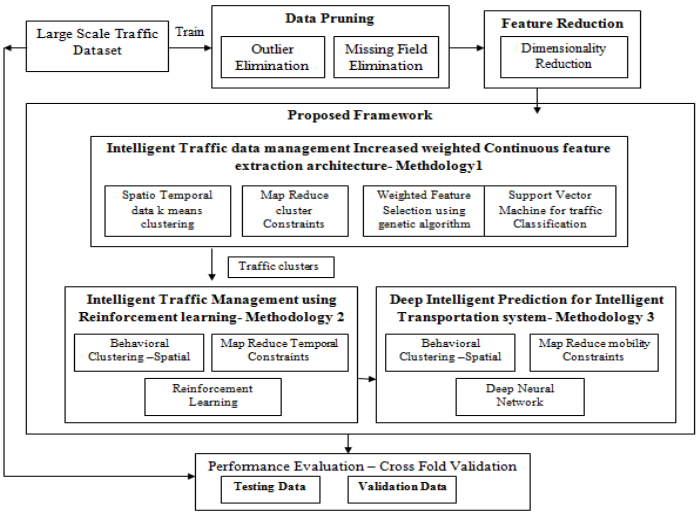

Figure A2.1: Proposed Framework of the Research

2.1. Smart Traffic Management System using Increased weighted Continuous feature extraction Algorithm

It is fully automated and is helpful in controlling traffic during emergencies. The proposed model uses supervised learning model to provide security against theft, set up sensors which help in identifying or send message regarding the traffic situation, and receive certain important message during emergency and pollution control on extracting the weight of the feature. It helps in calculating the average speed and also analyzes the passage of the vehicle. The model uses multiple SVM for prediction.

Traffic data is processed in terms of undirected graph G (V, E), where v represents the set of vertices and E represents the set of edges (Road Segments connecting two edges). The long edge is divided into shorter segments with length of 200meters. For POI and data trajectories, road network is categorized into clusters and estimate the potential information like traffic density, Vehicle moving speed and brake points etc for each cluster [11]. Top K road cluster are recommended to the passenger of the vehicle.

2.1.1. Pre-processing – Missing Value Prediction

Dataset is processed as verify whether variables or facts or records has lost. In order to fill the missed value, missing value prediction has been determined on employing K mean criterion. Data has been organized as graph containing edge and vertices to the preprocessed data information.

2.1.2. Spatiotemporal Clustering of Traffic Data

On adoption of graph clustering, optimal cluster will be obtained for dynamically changing traffic information's. Graph clustering generates the association matrix on based on the connectivity between vertices and the weights of the edges [12]. The association matrix for the traffic information in edge and vertices is as follows

$$\text{Association Matrix A} = = \begin{pmatrix} \operatorname{cov}(x,x) & \operatorname{cov}(x,y) & \operatorname{cov}(x,z) \\ \operatorname{cov}(y,x) & \operatorname{cov}(y,y) & \operatorname{cov}(y,z) \\ \operatorname{cov}(z,x) & \operatorname{cov}(z,y) & \operatorname{cov}(z,z) \end{pmatrix}$$

Matrix value is represented as $A_{ij} = \frac{1}{n-1}\sum_{m=1}^{n}\left(d1-\overline{X}_i\right)\left(dn-\overline{X}_j\right)+Xi$

The expansion process and inflation process will undergo different iteration until reaches the steady state. The clusters are generated from the matrix values.

2.1.3. Increased weighted Continuous Feature Extraction Algorithm

It extracts the normal and hidden features from the graph cluster using wrapper and filter method. Features are extracted mostly based on frequencies. For instance, features are extracted based on road types, traffic density, vehicle moving time and frequency of objects distributions near Point of interest. The features extracted as distinct variable. It helps to determine the free parking space and vehicle detection etc. However this mechanism, effectively computes the objects located in the circular area. Feature extraction is explained with multiple cases [14] .

- POI can be covered by a single cluster or multiple clusters.
- POI can be covered by a cluster for one time or multiple times.

The optimal feature selection has been carried out on model tuning during the learning process. It learns as single sequence learning.

2.2. Reinforcement Learning Algorithm

In this model, Adaptive traffic control system is a traffic management strategy used to control the traffic by facilitating the signals to instantaneously adjust to the present traffic demand. It has been implemented occurrence, replay and ideal mechanisms to improve the consistency of the algorithm. Q-learning using reinforcement has been designed in the precise form of the environment for selecting action represented in figure 2.1.

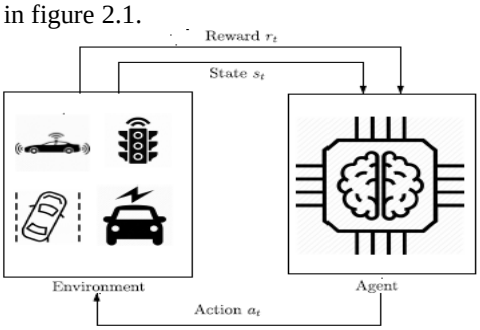

Figure 2.2: Architecture Deep reinforcement learning

2.2.1. Prediction

The Prediction model is employed to extracted features to control overcrowded traffic and for this the dynamic network with the linear signal arrangement using agent and rewards system on the set of states and actions for the effective policies. Reinforcement learning generally predicts its outcome only after number of repetitions (iterations) hence this method has chances of producing right decisions. It structurally comprises a hidden layer, along with a visible layer, where there are mutual connections between units in both the visible and hidden layers. The bias values for both the hidden layer and visible layer are updated. It automatically learns and infers the spatial dependency.

Algorithm 2.1 Reinforcement Learning
Input: Traffic data
Output: potential Cluster

Process

Initialize the network weight for features

Initialize the target function

$$P(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

The objective function in terms of various state is given by

$$P(x,y) = \alpha \frac{l_t(x,y)}{\sum_i l_t(x,y_i)} + (1-\alpha)\frac{f(x,y)}{\sum_i f(x,y)}$$

Where l_t is the location at specific time

Prediction is hierarchical and operates with spatiotemporal characteristics and location based service to predict potential road clusters for passengers [15].

2.2.2. Value Function:

The function computes the probability distribution over state-action pairs to explore all the states in a large and continuous space and store them in a table. It produces ranking in terms of probability than the other samples by applying a stochastic sampling with proportional prioritization or rank-based prioritization.

2.3. Deep Intelligent Prediction Network

The deep learning of the traffic data predicts the potential road clusters for passengers, traffic forecasting information as hierarchical model. Initially the underlying the traffic data is classified into clusters with extracting set of features on it. The deep behavioural network processes the traffic data in terms of spatiotemporal characteristics to generate the traffic forecasting information, vehicle detection, autonomous driving and driving behaviours.

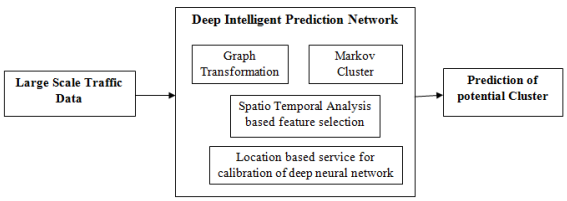

Figure 2.3: Architecture of Deep Intelligent Prediction Network

In addition, Vehicle detection and helps in autonomous driving behaviours. Moreover Hidden markov model adapted to system provides the hidden features such as road type and point of interest features. Architecture of the deep intelligent prediction framework is represented in the figure2.2.

2.3.1. Prediction using Deep Neural network

Deep Neural network can overcome this difficulty with a non-iterative training process. The features that is important for discrimination has been computed through normal and hidden layer. The prediction models are able to learn complicated functions and structures while irrelevant variations are suppressed.hih performance has been achieved in discovering the structure of large scale data.

Algorithm 2.2: Prediction of Traffic forecasting Clusters
Input: Traffic data
Output: potential Cluster
Process
For instance G = 1 to W do
Generate the instance a training data for Traffic data
For i = 1 to M do
Calculate class C = {V1, V2, V3...}
Where v1, v2 are input vectors of the road cluster
End for
Extract ()
Feature set = {V1(x, y), V2(x, y)...}
If Similarity difference between v1 and v2 > threshold
Initialize Random Cluster
Else
Optimize the parameter using Spatiotemporal Characteristics and weight

$$\text{Weight W} = V1\sum_{k=0}^{n}\binom{n}{k}x^k a^{n-k} + V1\sum_{k=0}^{n}\binom{n}{k}x^k a^{n-k} \ldots \text{Vn}$$

Predict ()

$$\text{P} = \frac{\text{W}\pm\sqrt{x^k a^{n-k}-4c}}{2\text{n}}$$

Recommend for specific Road segment ={ Available Parking space | vehicle Density | traffic Congestion }

It automatically learns and infers the spatial dependency. The sub layer which is responsible for reducing the size of the feature maps in the prediction operation.

3. Experimental Analysis

Experimental analysis is carried out with an Intel Core I3 processor with 2620 Processors (2.0 GHz) and 8 GB RAM and 1GB Hard disk using Dotnet programming on the real dataset which is described in the following section.

3.1. Dataset Description :

The proposed deep learning framework for intelligent transportation system has been analyzed on applying the data collected from the Caltrans Performance Measurement System (PeMS) database which consist of statewide freeway lane information from California. The traffic data are collected every 30 s from over 15 000 individual detectors are aggregated 5-min interval each for each detector station.The data taken for processing is 750 GB containing traffic information at various time periods and different location. Additionally GPS and other sensor information have been calibrated with location based service to obtain the POI.

3.2. Performance Measure

The performance metric considered for evaluation of the proposed intelligent transportation model consider the following

- Precision
- Average Execution Time

3.3. Performance Evaluation of the prediction Framework

The performance of the deep intelligent prediction framework has been obtained for different combinations and with minor fluctuations in the traffic data has been represented in Figure 3.1 describes the performance evaluation of the transportation models. The deep learning model prefers to consider both temporal and spatial correlations inherently on the location based services for proving the prediction for short-term passenger requirement to on-demand transport service for enjoying the intelligent transportation experience

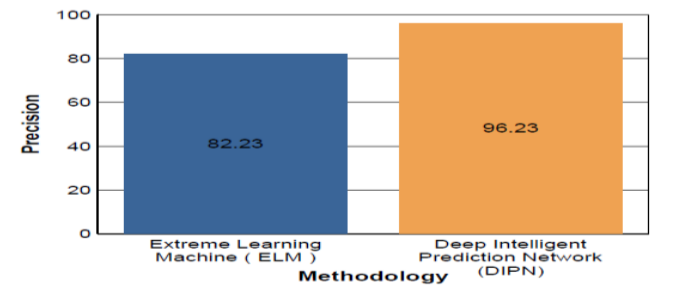

Figure 3.1: Performance analysis of Intelligent Transportation Prediction Model on Precision metric

The figure 3.2 represents the performance analysis of the intelligent transportation model on terms of average execution time. The traffic network spatial-temporal dependencies integration into deep learning models will reduce the computation complexities.

The model is represented in detail on their specification and Complex on their architecture. Trial and error method helps in extracting more knowledge thus is capable to decide whether to give positive or negative reward to the given set of state action pair. Hidden layer execute feature extraction, mathematical and other computations.

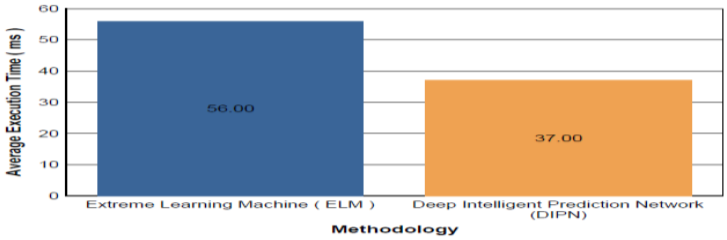

Figure 3.2: Performance analysis of Intelligent Transportation Prediction Model on Average Execution Time

Other factors considered for enhancing the performance includes land use and origin-destination data. Generalization error model has been considered. The table 4.1 depicts the performance evaluation of prediction models.

Table 4.1: Performance evaluation of Intelligent Transportation Prediction models

Technique	Precision in percentage	Average Execution Time in ms
Extreme Learning Machine (ELM)	82.23	56
Increased Weighted Continuous Feature Extraction using SVM	92.56	46
Deep Reinforcement Learning	94.25	41
Deep Intelligent Prediction Network (DIPN)	96.23	37

Finally, the efficiency of the proposed model has been improved when compared with state of art approaches with satisfactory results.

Organization of the thesis

The organization of the thesis is sectioned into following chapters with an emphasis contribution made on the modelling a framework to the devise a Big Data Cloud Driven Computational Framework for Intelligent Transportation System for traffic data exploring in the volume and velocity.

Chapter 1- Introduction

In this chapter, basic introduction of the Transportation Model, Big Data Architecture for scaling the transportation data, Application areas of transportation models and traffic control schemes. Further we discuss the problem definition, motivation and contribution of the research

Chapter 2- Literature review

In this chapter, detailed background analysis of traffic Data management using Big Data Fusing models, Heterogeneous Data Learning approaches and Location based Recommendation System has been carried out on various category and on the computed properties of the data.

Chapter 3- Smart Traffic Management System using Increased weighted Continuous feature extraction Algorithm

In this chapter, we model an increased weighted Continuous feature extraction Algorithm on traffic data towards controlling and providing security against theft through inclusion of multiple SVM.

Chapter 4- Reinforcement learning for controlling traffic data

In this chapter, we model an adaptive traffic control system to control the traffic by facilitating the signals to instantaneously adjust to the present traffic demand. It has been implemented occurrence, replay and ideal mechanisms to improve the consistency of the algorithm.

Chapter 5- Deep intelligent prediction Framework

In this chapter, deep intelligent prediction Framework has been designed towards prediction of the potential cluster for passengers on traffic information based on based spatiotemporal characteristics and location based services

Chapter 6- Results and Discussion

In this chapter the result are analysed using large dataset, computed the classification performance against various mechanism using metric.

Chapter 7- Conclusion

The conclusion of the work has been described with hard step taken to model a new solution to control the traffic data using big data architecture with several highlights along accurate results and it is journal publication is also discussed.

Conclusion

We designed and implemented a new framework named Big Data Cloud Driven Computational Framework for Intelligent Transportation System on employing as deep intelligent network, Increase weighted continuous feature extraction, reinforcement learning towards prediction of the potential cluster for passengers on traffic information. The prediction model conducted to cluster and classify the traffic information based spatiotemporal characteristics and location based services as big data paradigm. The feature set is designed based on hidden markov model and normal feature extraction models. It reflects the various characteristics with regards to various factors.

The clustered information enables effective utilization of the data. However there have been some efforts made to system to recommend the passengers with parking spaces, road based traffic density and time based traffic density and finally it aids in autonomous driving of the vehicle. The performance of the proposed framework has been evaluated through extensive experiment based on traffic data. On analysis of Experimental results, the flexibility and scalability of the model has found to be improved over state of art approaches.

References

[1] Y. Liu, C. Kang, S. Gao, Y. Xiao, and Y. Tian, "Understanding intra-urban trip patterns from taxi trajectory data," Journal of Geographical Systems, vol. 14, no. 4, pp. 463–483, 2012.

[2] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-Drive: Enhancing driving directions with taxi drivers' intelligence," IEEE Transaction in Knowledge and Data engineering, Vol.3, no.9, pp: 201-2013, 2008

[3] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-Finder: A recommender system for finding passengers and vacant taxis," IEEE Transactions Knowledge and Data Engineering., vol. 25, no. 10, pp. 2390–2403, Oct. 2013.Eng., vol. 25, no. 1, pp. 220–232, Jan. 2013.

[4] G. Chen, B. Chen, and Y. Yu, "Mining frequent trajectory patterns from GPS tracks," in Proceedings in IEEE International Conference Computer Intelligent Software Engineering. 2010, pp. 1–6.

[5] Y. Zhu, Y. Wu, and B. Li, "Trajectory improves data delivery in urban vehicular networks," IEEE Transaction on Parallel and Distributed system., vol. 25, no. 4, pp. 1089–1100, Apr. 2014.

[33] D. Zhang, T. He, Y. Liu, and J. A. Stankovic, "CallCab: A unified recommendation system for carpooling and regular taxicab services," in Proceedings in IEEE International Conference on Big Data, 2013, pp. 439–447.

[6] Jin, F., Sun, S.: 'Neural network multitask learning for traffic flow forecasting'. IEEE International Joint Conference on Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1 June 2008, pp. 1897–1901

[7] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," IEEE Trans. Syst.Man Cybern. Part B (Cybernetics), vol.42, no. 2, pp. 513–529, 2012.

[8] D. S. Yeung, W. W. Y. Ng, D. Wang, E. C. C. Tsang, and X.-Z. Wang, "Localized generalization error model and its application to architecture selection for radial basis function neural network,"

[9] El-Tantawy, S., Abdulhai, B., Abdelgawad, H.: 'Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto', IEEE Transaction on Intelligent Transporation System., 2013, 14, (3), pp. 1140–1150

[10] Liu, L., Chen, R.C.: 'A MRT daily passenger flow prediction model with different combinations of influential factors'. 2017 31st Int. Conf. Advanced Information Networking and Applications Workshops (WAINA), Taipei, Taiwan, 27 March 2017, pp. 601–605

[11] R. Wang, S. Kwong, and D. D. Wang, "An analysis of ELM approximate error based on random weight matrix," International Journal of Uncertainty Fuzziness Knowledge Based System., vol. 21, pp. 1–12, 2013.

[12] Xu, J., Rahmatizadeh, R., Bölöni, L., et al.: 'Real-time prediction of taxi demand using recurrent neural networks', IEEE Trans. Intelligent TransportationSystem,. 2017, pp. 1–10

[13] X.-Z. Wang, Q.-Y. Shao, Q. Miao, and J.-H. Zhai, "Architecture selection for networks trained with extreme learning machine using localized generalization error model," Neurocomputing, vol. 102, pp. 3–9, 2013.

[14] S. Van Dongen, "Graph clustering via a discrete uncoupling process," Journal of Matrix Analytical. Applications, vol. 30, no. 1, pp. 121–141, 2008.

[15] Y. J. Li, Y. Li, J. H. Zhai, and S. Shiu, "RTS game strategy evaluation using extreme learning machine," Soft Computing., vol. 16, no. 9, pp. 1627–1637, 2012.